

AD-A033 341

STANFORD UNIV CALIF DEPT OF OPERATIONS RESEARCH
ESTIMATING THE TRUE PROBABILITY OF CORRECT SELECTION FOR LOCATI--ETC(U)
JUN 76 I OLKIN, M SOBEL, Y L TONG

F/G 12/2

N00014-75-C-0561

NL

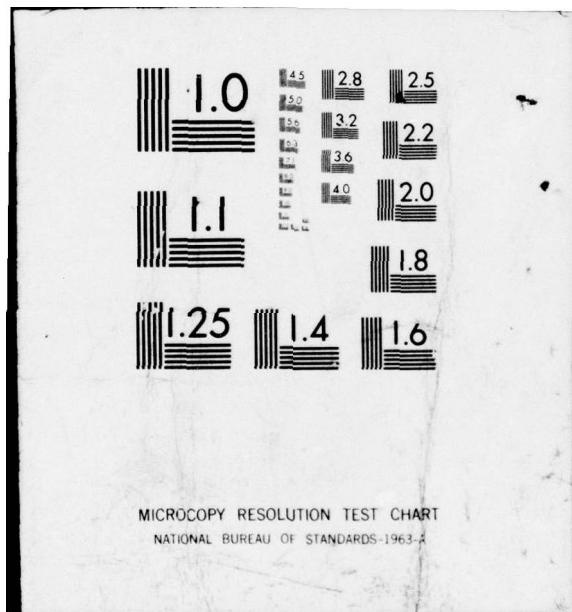
UNCLASSIFIED

TR-174

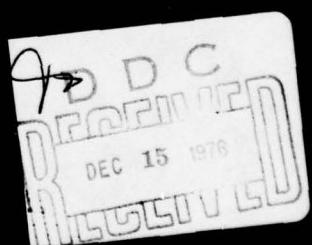
| OF |
AD
A033341

END

DATE
FILMED
2-77



ADA033341



ESTIMATING THE TRUE PROBABILITY OF CORRECT SELECTION
FOR LOCATION AND SCALE PARAMETER FAMILIES

by

I. OLKIN, M. SOBEL and Y.L. TONG

TECHNICAL REPORT NO. 174

JUNE 28, 1976

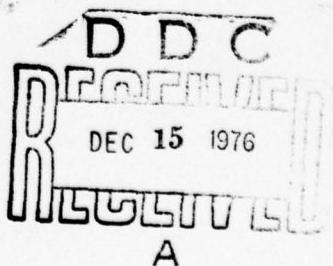
SUPPORTED BY THE ARMY AND NAVY
UNDER CONTRACT N00014-75-C-0561 (NR-042-002)
WITH THE OFFICE OF NAVAL RESEARCH

Gerald J. Lieberman, Project Director

Reproduction in Whole or in Part is Permitted
for any Purpose of the United States Government
Approved for public release; distribution unlimited.



DEPARTMENT OF OPERATIONS RESEARCH
AND
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



ESTIMATING THE TRUE PROBABILITY OF CORRECT SELECTION
FOR LOCATION AND SCALE PARAMETER FAMILIES

BY

I. Olkin, M. Sobel and Y. L. Tong
Stanford University; University of California,
Santa Barbara; University of Nebraska

Abstract

The usual formulation of the problem of selecting the best of k populations has been studied from the point of view of designing an experiment that will guarantee a certain probability of making a correct selection (PCS). However, in retrospective studies the sample has already been taken and we wish to estimate the probability of making a correct selection using this data. The present study deals with obtaining a class of upper and lower bounds for the true PCS and its estimator using the methods of majorization. We also study the large sample behavior of the estimator of the PCS, which turns out to be asymptotically normal.

a

1. Introduction.

Consider an independent observation, X_{ij} , from k populations $F(x, \theta_i)$, $i = 1, \dots, k$, $j = 1, \dots, n$. The experimenter wishes to select the population associated with the largest parameter $\theta_{[k]}$. To do this we define an appropriate statistic

$$Y_i = Y(X_{i1}, \dots, X_{in}) ,$$

and use the decision rule "designate the population corresponding to the largest Y_i value as the one with the largest θ -value." If the c.d.f., $G(y, \theta_i)$, of Y_i has the form $G(y - \theta_i)$, then the probability of a correct selection (PCS) is

$$(1.1) \quad \alpha(\delta_1, \dots, \delta_{k-1}) = \int_{-\infty}^{\infty} \left[\prod_{l=1}^{k-1} G(y + \delta_l) \right] dG(y) ,$$

where $\theta_{[1]} \leq \dots \leq \theta_{[k]}$, $\delta_i = \theta_{[k]} - \theta_{[i]}$.

In the usual methods employing the "indifference zone" approach, as introduced into ranking and selection by Bechhofer (1954), we determine the smallest fixed sample size n needed in order to satisfy a preassigned probability requirement. This led to a consideration of least favorable configuration, which frequently takes the form $\delta_1 = \dots = \delta_{k-1} \equiv \delta_0$ and suggests an inequality for $\alpha(\delta_1, \dots, \delta_{k-1})$, where we "move downhill" in the sense of majorization to the case of equal δ_i -values, regardless of whether the common value δ_0 is specified or not. Because this integral is often difficult to evaluate (especially in high dimensions), we are interested in both upper and lower bounds

that hold for any true configuration, without regard to any least favorable configuration.

In the present paper we consider the problem of estimating the true PCS by using information provided by the sample. This is extremely important for those problems in which we are carrying out a retrospective study -- rather than designing an experiment.

If Y is a consistent estimator of θ , then we may let

$y_{[1]} \leq \dots \leq y_{[k]}$ be the ordered Y -values, and define

$$\hat{\delta}_i = y_{[k]} - y_{[i]} , \quad i = 1, \dots, k-1 .$$

Then $\hat{\delta} = (\hat{\delta}_1, \dots, \hat{\delta}_{k-1})$ will be a consistent estimator of δ , and $\alpha(\hat{\delta}_1, \dots, \hat{\delta}_{k-1})$, will be a consistent estimator of $\alpha(\delta_1, \dots, \delta_{k-1})$.

As a consequence, we wish to further study the behavior of the integral $\alpha(\delta_1, \dots, \delta_{k-1})$, or $\alpha(\hat{\delta}_1, \dots, \hat{\delta}_{k-1})$, as a function of the configuration of the parameters. In particular, we are interested in the following question: how does the PCS behave if some of the δ -values are replaced by their average? This problem is studied in the next section.

Remark: The above development was based on a location parameter family. However, a similar analysis holds for scale parameter families. In this case the true PCS (1.1) is replaced by

$$(1.2) \quad \alpha(\delta_1, \dots, \delta_{k-1}) = \int_{-\infty}^{\infty} \left[\prod_{i=1}^{k-1} G(\delta_i y) \right] dG(y) ,$$

where in the case of the largest scale parameter we define $\delta_i = \theta_{[k]}/\theta_{[i]}$ ($i = 1, 2, \dots, k-1$) and hence $\delta_1 \geq \dots \geq \delta_{k-1} \geq 1$. For the case of the smallest scale parameter we define $\delta_{k-i+1} = \theta_{[i]}/\theta_{[1]}$ and then (1.2) and the above inequalities on δ_i still hold.

2. The Main Bounds.

The key motivating point is that under certain circumstances we can approximate $\alpha(\delta)$ by

$$\alpha(\delta_1, \dots, \delta_p) \cong \alpha(\underbrace{\delta_1, \dots, \delta_1}_{p_1}, \underbrace{\delta_2, \dots, \delta_2}_{p_2}, \dots, \underbrace{\delta_r, \dots, \delta_r}_{p_r}),$$

where $\sum p_i = p \equiv k-1$ and the $\bar{\delta}_i$'s are certain averages of the δ 's.

If the approximation sign is an inequality sign, then we are "moving downhill" by averaging and further averaging will move us more downhill. This suggests the use of majorization, and we show how this arises.

We now develop the main result. For any fixed $r \leq p$, let a_1, \dots, a_r be distinct numbers satisfying

$$0 \equiv a_0 < a_1 < a_2 < \dots < a_r \equiv p,$$

$p_v = a_v - a_{v-1}$, $v = 1, \dots, r$, and define the partition

$$\delta(A) = (\underbrace{\delta_1, \dots, \delta_{a_1}}_{p_1}, \underbrace{\delta_{a_1+1}, \dots, \delta_{a_2}}_{p_2}, \dots, \underbrace{\delta_{a_{r-1}+1}, \dots, \delta_p}_{p_r}).$$

We also define the means within each cluster:

$$\bar{\delta}_v = \frac{\delta_{a_{v-1}+1} + \dots + \delta_{a_v}}{a_v - a_{v-1}}, \quad v = 1, \dots, r,$$

and generate a vector of "averaged" δ -values:

$$\begin{aligned}\bar{\delta}(A) &\equiv (\bar{\delta}_1(A), \dots, \bar{\delta}_p(A)) \\ &= (\underbrace{\bar{\delta}_1, \dots, \bar{\delta}_1}_{p_1}, \underbrace{\bar{\delta}_2, \dots, \bar{\delta}_2}_{p_2}, \dots, \underbrace{\bar{\delta}_r, \dots, \bar{\delta}_r}_{p_r}),\end{aligned}$$

where A denotes the set $A = \{a_1, \dots, a_r\}$.

Let $A \subset B$ mean that B is a refinement of A . A question that arises immediately is whether we can compare the α -values of two partitions A and B if $A \subset B$?

Theorem 1. If $\log G(y)$ is concave, and $A \subset B$, then

$$(2.1) \quad \alpha(\delta_1, \dots, \delta_p) \leq \alpha(\bar{\delta}_1(B), \dots, \bar{\delta}_p(B)) \leq \alpha(\bar{\delta}_1(A), \dots, \bar{\delta}_p(A)).$$

The proof of this result is given in the next section. We note that the α 's in (2.1), though written as a single integral, are actually p -fold integrals. However, we would also like to have bounds that reduce the dimensionality. To accomplish this we consider the following integral. For any partition A , define

$$(2.2) \quad \beta(\delta_1(A), \dots, \delta_p(A))$$

$$= \left[\int_{-\infty}^{\infty} \prod_{i=1}^{a_1} G(y+\delta_i) dG(y) \right] \left[\int_{-\infty}^{\infty} \prod_{i=a_1+1}^{a_2} G(y+\delta_i) dG(y) \right]$$

$$\cdots \left[\int_{-\infty}^{\infty} \prod_{i=a_{r-1}+1}^p G(y+\delta_i) dG(y) \right].$$

Here the reduction in dimensionality is in terms of the size of the intervals $a_v - a_{v-1}$ ($v = 1, 2, \dots, r$). Note that when $r = p$, then $\beta(\delta_1, \dots, \delta_p) = \alpha(\delta_1, \dots, \delta_p)$.

Theorem 2. If $A \subset B$, then

$$(2.3) \quad \beta(\delta_1(B), \dots, \delta_p(B)) \leq \beta(\delta_1(A), \dots, \delta_p(A)) \leq \alpha(\delta_1, \dots, \delta_p).$$

Of particular interest are the extremes obtained by the partition $A = \{1, \dots, p\}$, in which case we obtain

$$(2.4) \quad \prod_{i=1}^p \int_{-\infty}^{\infty} G(y+\delta_i) dG(y) \leq \alpha(\delta_1, \dots, \delta_p) = \int_{-\infty}^{\infty} \prod_{i=1}^p G(y+\delta_i) dG(y).$$

Remarks: (i) We should note that Theorem 2 is due to Kimball (1951).
(ii) If G is a normal distribution, then the right-hand side of (2.3) can be regarded as the integral of a multivariate normal distribution with variances 1 and correlations $\rho = 1/2$. This integral has been considered by Milton (1970, 1972). If G is normal then the left-hand side of (2.4) is easily shown to be $\prod_{i=1}^p G(\delta_i/\sqrt{2})$, so that no integration is required.

(iii) If $G(y, \theta) = G(y/\theta)$, with $G(0) = 0$, is a scale parameter family, then we define geometric means instead of arithmetic means.

For any partition A we have

$$\bar{\delta}_1 = \left(\prod_{i=1}^{a_1} \delta_i \right)^{1/a_1}, \quad \bar{\delta}_2 = \left(\prod_{i=a_1+1}^{a_2} \delta_i \right)^{1/(a_2-a_1)}, \\ \dots, \quad \bar{\delta}_r = \left(\prod_{i=a_{r-1}+1}^p \delta_i \right)^{1/(p-a_{r-1})}.$$

The results of Theorems 1 and 2 hold with the obvious changes.

3. Proof of Inequalities via Majorization.

We first review several facts concerning the partial ordering majorization. A vector u_1, \dots, u_p is said to majorize a vector v_1, \dots, v_p , written $u > v$, if, upon a rearrangement with $u_1 \geq \dots \geq u_p$, $v_1 \geq \dots \geq v_p$, we have $\sum_1^m u_i \geq \sum_1^m v_i$, $m = 1, \dots, p-1$, $\sum_1^p u_i = \sum_1^p v_i$. The class of (differentiable) order preserving functions $f(x_1, \dots, x_p)$ are called Schur functions, or Schur-convex functions, and are defined by the property

$$\left(\frac{\partial f(x)}{\partial x_i} - \frac{\partial f(x)}{\partial x_j} \right) (x_i - x_j) \geq 0 \quad \text{for all } i, j = 1, \dots, p.$$

(For the Schur-concave functions we merely turn the inequality around.)

Then one characterization of the Schur functions is through the fact that $u > v$ if and only if $f(u) \geq f(v)$ for all Schur-convex functions.

Finally, there is another characterization in terms of doubly-stochastic matrices $P = (p_{ij})$, $p_{ij} \geq 0$, $\sum_j p_{ij} = \sum_i p_{ij} = 1$, namely, $u > v$ if and only if $v = uP$ for some doubly stochastic matrix P . This says that one "moves downhill" from u to v (in the sense of majorization) by taking a convex combination of elements of u .

We need the following fact.

Fact. [Marshall and Olkin (1974)]. If $\log G(y)$ is concave, then

$\prod_1^p G(y+\delta_i)$ is Schur-concave. Consequently, $\alpha(\delta) = \int_{-\infty}^{\infty} \prod_1^p G(y+\delta_i) dG(y)$ is Schur-concave.

The proof of Theorem 1 is a direct consequence of the above facts. Recall that in (2.1) we have that if $A \subset B$ then

$$\alpha(\delta) \leq \alpha(\delta(B)) \leq \alpha(\delta(A)).$$

We need only prove the left-hand inequality since the right-hand inequality is obtained by a repeated argument where $\alpha(\delta(B))$ is used in place of $\alpha(\delta)$ and $\alpha(\delta(A))$ is used in place of $\alpha(\delta(B))$.

It is immediate that $\delta > \delta(B)$ since $\delta(B) = \delta P$, where $P = \text{block diag}(P_1, \dots, P_r)$ is a doubly-stochastic matrices with blocks P_v on the diagonal consisting of $1/(a_v - a_{v-1})$. This majorization plus the Schur concavity proves Theorem 1.

Theorem 2 follows from an inequality of Kimball [1951] and to apply this it is sufficient that the distribution functions $G(y+\delta_i)$ all be positive and monotone in the same direction for all real y .

4. Approximations.

We now make use of the bounds to approximate $\alpha(\delta)$. Although $\alpha(\delta)$ can be computed exactly with the help of quadrature methods, the necessary programming is not always available and the computations may be costly. Thus, it is desirable to be able to approximate the integrals.

As an illustration of the bounds, consider a numerical example for the case of normality with common $\sigma = 1$, and the location parameter model with $\delta = (3.2, 2.7, 2.5, 1.9, 1.7)$; we wish to evaluate $\alpha(\delta)$.

The exact value is $\alpha = 0.8016$. To compute our smallest lower bound

note that $\int G(y+\delta) dG(y) = \Phi(\delta/\sqrt{2})$, so that we obtain $\prod_1^5 \Phi(\delta_i/\sqrt{2}) = 0.7443$. The largest upper bound is $\int \Phi^5(y+\delta) d\Phi(y) = 0.8483$, obtained from Milton (1970).

We now consider some refinements. Suppose we use $A = \{1, 3, 5\}$ so that we cluster into 3 groups $\{3.2\}, \{2.7, 2.5\}, \{1.9, 1.7\}$, then the lower bound $B(\delta(A))$ in (2.2) is given by

$$(4.1) \quad \int \Phi(y+3.2) d\Phi(y) + \int \Phi(y+2.7) \Phi(y+2.5) d\Phi(y) \\ + \int \Phi(y+1.9) \Phi(y+1.7) d\Phi(y),$$

which may be obtained from the Milton (1970) tables to be 0.7698.

But, if we now approximate (4.1) by

$$(4.2) \quad \int \Phi(y+3.2) d\Phi(y) \int \Phi^2(y+2.6) d\Phi(y) \int \Phi^2(y+1.8) d\Phi(y),$$

then we obtain the result 0.7717.

If we use fewer clusters, namely, $A = \{3,5\}$, yielding $\{3.2, 2.7, 2.5\} \{1.9, 1.7\}$ then we obtain the approximation

$$(4.3) \quad \int \phi^3(y+2.8) d\Phi(y) \int \phi^2(y+1.8) d\Phi(y)$$

to yield a value of 0.7802.

It is intuitively clear that if we can take advantage of the close proximity of any two $\hat{\delta}$ values, then the worst case for getting close bounds and approximations may be the equally spaced case. We have, therefore, attached at the end of this paper a table (Table 1) for the normal distribution under equally spaced configurations with a common spacing d . The table gives the numerical values of the integral

$$\int \prod_{i=1}^p \Phi(y+id) d\Phi(y) = P[Z_i \leq id/\sqrt{2}, i = 1, \dots, p] \text{ for}$$

$d = 0.00(0.02)0.20(0.10)4.00$ and $k = 2(1)10, 12, 15$ and 20;

where Φ is the standard normal cdf.

5. Limiting Distribution of the Estimator of the PCS for the Normal Location Parameter Model.

In this section we obtain the asymptotic distribution of the estimator, $\hat{\alpha}(\hat{\delta})$, of the true PCS $\alpha(\delta)$ when the underlying distribution is normal. Although results for the location parameter case are obtained, similar results may be obtained for scale parameter families.

Let $Y = (Y_1, \dots, Y_k)$ and $\tilde{Y} = (Y_{[1]}, \dots, Y_{[k]})$, where $Y_{[1]} \leq \dots \leq Y_{[k]}$ are the ordered Y -values. Similarly, we define

$\theta = (\theta_1, \dots, \theta_k)$ and $\tilde{\theta} = (\theta_{[1]}, \dots, \theta_{[k]})$. Finally we let $\Omega = \{\theta : \theta_{[1]} < \dots < \theta_{[k]}\}$ with strict inequalities between θ -values. For any two vectors u and v we write $u \leq v$ to mean $u_i \leq v_i$ for all i .

Now suppose $L(\sqrt{n}(Y_i - \theta_i)) \xrightarrow{d} N(0, 1)$. We first obtain the asymptotic distribution of \tilde{Y} , then of estimators, \hat{s}_i , of s , and finally of $\alpha(\hat{s})$. In the following development σ is assumed to be known, and we take it to be unity.

Theorem 3. As $n \rightarrow \infty$, $L(\sqrt{n}(\tilde{Y} - \tilde{\theta})) \xrightarrow{d} N(0, I_k)$ for every $\tilde{\theta} \in \Omega$, where I_k is the identity matrix of order k .

Proof. The result follows by bounding the right-hand side of

$$(5.1) \quad \begin{aligned} |P_{\tilde{\theta}}\{\sqrt{n}(\tilde{Y} - \tilde{\theta}) \leq z\} - \prod_1^k \Phi(z_i)| \\ \leq |P_{\tilde{\theta}}\{\sqrt{n}(Y - \tilde{\theta}) \leq z, Y = \tilde{Y}\} - P_{\tilde{\theta}}\{\sqrt{n}(Y - \tilde{\theta}) \leq z\}| \\ + P_{\tilde{\theta}}\{\sqrt{n}(\tilde{Y} - \tilde{\theta}) \leq z, Y \neq \tilde{Y}\} + |P_{\tilde{\theta}}\{\sqrt{n}(Y - \tilde{\theta}) \leq z\} - \prod_1^k \Phi(z_i)|, \end{aligned}$$

for every $z = (z_1, \dots, z_k)$ and every $\tilde{\theta} \in \Omega$.

Let $\epsilon_1 > 0$, $\epsilon_2 > 0$ be arbitrary fixed real numbers such that $2\epsilon_1 \leq \min_i(\theta_{[i]} - \theta_{[i-1]})$. Since \tilde{Y} is a consistent estimator of $\tilde{\theta}$ and the Y_i 's are independent, it follows that for every $\tilde{\theta} \in \Omega$ and every $z = (z_1, \dots, z_k)$ there exist numbers $N_1 \equiv (\epsilon_1, \epsilon_2)$ and $N_2 \equiv N_2(\epsilon_2, z)$ such that for $n > N_1$, $P_{\tilde{\theta}}\{Y = \tilde{Y}\} \geq P_{\tilde{\theta}}\{|Y_i - \tilde{\theta}_i| \leq \epsilon_1, i = 1, \dots, k\} \geq 1 - \epsilon_2$, and for $n > N_2$, $|P_{\tilde{\theta}}\{\sqrt{n}(Y - \tilde{\theta}) \leq z\} - \prod_1^k \Phi(z_i)| \leq \epsilon_2$.

Here ϵ_1 is small enough so that $|Y_i - \tilde{\theta}_i| \leq \epsilon_1$ implies that $Y = \tilde{Y}$, so that the first difference on the right side of (5.1) is zero. Then for $n > \max\{N_1, N_2\}$, (5.1) is bounded by $2\epsilon_2$, which completes the proof. ||

As a consequence of Theorem 3, we have the

Corollary 4. For every $\tilde{\theta} \in \Omega$, $\mathcal{L}(\sqrt{n}(\hat{\delta} - \theta))$ is asymptotically normal with mean vector zero, variances 2 and covariances 1.

Now consider the PCS for the normal case

$$\alpha(\delta) = a(\delta_1, \dots, \delta_p) = \int_{-\infty}^{\infty} \prod_{i=1}^p \Phi(t + \sqrt{n}\delta_i) d\Phi(t),$$

and its estimator,

$$\hat{\alpha}(\hat{\delta}) = a(\hat{\delta}_1, \dots, \hat{\delta}_p) = \int_{-\infty}^{\infty} \prod_{i=1}^p \Phi(t + \sqrt{n}\hat{\delta}_i) d\Phi(t).$$

Theorem 4. $\mathcal{L}(\sqrt{n}[a(\hat{\delta}) - a(\delta)]) \xrightarrow{d} N(0, v_n)$, where $v_n(\delta) = \sum_{i=1}^p d_i^2 + (\sum_{i=1}^p d_i)^2$, and

$$(5.2) \quad d_i = \frac{\sqrt{n}}{\sqrt{2}} \Phi\left(\frac{\sqrt{n}\delta_i}{\sqrt{2}}\right) \int_{-\infty}^{\infty} \prod_{j \neq i} \Phi\left(\frac{t}{\sqrt{2}} + \sqrt{n}\delta_j - \frac{\sqrt{n}\delta_i}{2}\right) d\Phi(t).$$

Proof. A Taylor expansion about $\delta = \delta_0 = (\delta_{10}, \dots, \delta_{p0})$ yields

$$\begin{aligned} \sqrt{n}(a(\hat{\delta}) - a(\delta_0)) &= \sqrt{n} \sum_{i=1}^p d_i (\hat{\delta}_i - \delta_{i0}) + \left(\frac{\sqrt{n}}{2} \sum_i (\hat{\delta}_i - \delta_{i0})^2 \left. \frac{\partial^2 a(\delta)}{\partial \delta_i^2} \right|_{\delta=\delta_0} \right. \\ &\quad \left. + \frac{\sqrt{n}}{2} \sum_{i \neq j} (\hat{\delta}_i - \delta_{i0})(\hat{\delta}_j - \delta_{j0}) \left. \frac{\partial^2 a(\delta)}{\partial \delta_i \partial \delta_j} \right|_{\delta=\delta_0} \right) \\ &\approx w_1(n) + w_2(n), \end{aligned}$$

where $\hat{\delta}_i \leq \delta_{i0} \leq \delta_i$. Corollary 4 and the boundedness condition

$\left| \frac{\partial a(\delta)}{\partial \delta_i} \right| \leq c/\delta_i$ implies that $L(w_1(n)) \xrightarrow{d} N(0, v_n)$. Consequently,

the proof will be complete if we show that $\text{plim } w_2(n) = 0$. To see this, we use the fact that $x\varphi(x)$ and $x^2\varphi(x)$ are bounded, so that there exist positive constants c_1, c_2, c_3 such that

$$\left| \frac{\partial^2 a(\delta)}{\partial \delta_i^2} \right| \leq \frac{c_1}{\delta_i^2}, \quad \left| \frac{\partial^2 a(\delta)}{\partial \delta_i \partial \delta_j} \right| \leq \frac{c_2}{\delta_i} + \frac{c_3}{\delta_i^2}.$$

If we define

$$U(n) = \frac{\sqrt{n}}{2} \sum_1^p (\hat{\delta}_i - \delta_{i0})^2 \frac{c_1}{\delta^{*2}} + \frac{\sqrt{n}}{2} \sum_{i \neq j} |(\hat{\delta}_i - \delta_{i0})(\hat{\delta}_j - \delta_{j0})| \left(\frac{c_2}{\delta^*} + \frac{c_3}{\delta^{*2}} \right),$$

then $\text{plim } U(n) = 0$. Consequently, if $A = \{\hat{\delta}_i > \delta^*, i = 1, \dots, p\}$, for any $\delta^* < \delta_p$, it follows that

$$(5.3) \quad P(|w_2(n)| > \epsilon) = P(|w_2(n)| > \epsilon, A) + P(|w_2(n)| > \epsilon, \bar{A}) \\ \leq P(U_n > \epsilon) + P(\bar{A}).$$

But for fixed $\epsilon > 0$ and sufficiently large n , each term on the right-hand side of (5.3) is $\leq \epsilon$. ||

The asymptotic variance v_n defined in Theorem 4 is quite complicated. We may use $\hat{\delta}_i$ in $v_n(\delta)$ to yield a confidence interval for the PCS, $a(\delta)$, namely,

$$\frac{\sqrt{n} (a(\delta) - a(\hat{\delta}))}{\sqrt{v_n(\hat{\delta})}} \sim N(0, 1).$$

However, we may also wish to approximate the asymptotic variance $v_n(\delta)$ or its estimate $v_n(\hat{\delta})$. We again use majorization to obtain some bounds. To do this we require the following facts.

Fact 1. For fixed i , and any z , the function

$$\prod_{j \neq i} \Phi\left(\frac{t}{\sqrt{2}} + \sqrt{n} \delta_j - \frac{\sqrt{n} \delta_i}{2}\right)$$

is a Schur-concave function in $(\delta_1, \dots, \delta_{i-1}, \delta_{i+1}, \dots, \delta_p)$.

This follows from the result of Section 3.

Fact 2. For fixed i and $\delta_1 \geq \dots \geq \delta_p$,

$$\tilde{\delta} \equiv \left(\delta_1 - \frac{\delta_i}{2}, \dots, \delta_{i-1} - \frac{\delta_i}{2}, \delta_{i+1} - \frac{\delta_i}{2}, \dots, \delta_p - \frac{\delta_i}{2} \right)$$

$$\succ \left(\frac{p}{p-1} \bar{\delta} - \frac{(p+1)}{2(p-1)} \delta_i, \dots, \frac{p}{p-1} \bar{\delta} - \frac{(p+1)}{2(p-1)} \delta_i \right) \equiv \delta_L,$$

$$\text{where } \bar{\delta} = \sum_1^p \delta_i / p.$$

To prove this result we note that the elements of δ_L are means of the elements in $\tilde{\delta}$.

Fact 3. For fixed i and $\delta_1 \geq \dots \geq \delta_p$,

$$\tilde{\delta} = \left(\delta_1 - \frac{\delta_i}{2}, \dots, \delta_{i-1} - \frac{\delta_i}{2}, \delta_{i+1} - \frac{\delta_i}{2}, \dots, \delta_p - \frac{\delta_i}{2} \right)$$

$$\prec \left(p\bar{\delta} - (p-2)\delta_p - \frac{3}{2}\delta_1, \delta_p - \frac{\delta_1}{2}, \dots, \delta_p - \frac{\delta_1}{2} \right) \equiv \delta^U.$$

To prove this majorization we first note that $p\bar{\delta} - (p-2)\delta_p - \frac{3}{2}\delta_1 \geq \delta_p - \delta_1/2$. Further, the sum of all the elements in each vector is equal to $p\bar{\delta} - \frac{1}{2}(p+1)\delta_1$. Consequently, we need show that for $m = 1, \dots, p-1$,

$$m(\delta_p - \frac{\delta_1}{2}) \leq \sum_{\substack{j=m+1 \\ j \neq i}}^p \delta_j - \frac{m\delta_i}{2} .$$

But this inequality is equivalent to

$$\sum_{\substack{j=m+1 \\ j \neq i}}^p \delta_j / (m-1) \geq \delta_p - (\delta_1 - \delta_i)/2 ,$$

which clearly holds.

As a consequence of the majorization

$$\delta_L \prec \tilde{\delta} \prec \delta^U ,$$

we have that for every z and every i

$$\prod_{j \neq i} \Phi \left(\frac{t}{\sqrt{2}} + \sqrt{n} \delta_j - \sqrt{n} \frac{\delta_i}{2} \right)$$

$$\begin{cases} \leq \Phi^{p-1} \left(\frac{t}{\sqrt{2}} + \frac{\sqrt{n} p\bar{\delta}}{p-1} - \frac{\sqrt{n}(p+1)}{2(p-1)} \delta_i \right) \equiv D_2(t) , \\ \geq \Phi \left(\frac{t}{\sqrt{2}} + \sqrt{n} p\bar{\delta} - (p-2)\sqrt{n} \delta_p - \frac{3}{2}\sqrt{n} \delta_1 \right) \Phi^{p-2} \left(\frac{t}{\sqrt{2}} + \sqrt{n} \delta_p - \frac{\sqrt{n} \delta_1}{2} \right) \equiv D_1(t) . \end{cases}$$

Remark. The upper bound is a monotone decreasing function of δ_i , and the lower bound is a monotone increasing function of δ_p . Consequently, if $\delta_i \geq \delta^*$, we may obtain the bounds where δ^* is inserted in place of δ_i in $D_2(t)$ and in place of δ_p in $D_1(t)$. Call these functions $\tilde{D}_j(t)$, $j = 1, 2$. From the definition of v_n , we see that

$$\frac{n}{4\pi} H(\delta) [\int \tilde{D}_1(t) d\Phi(t)]^2 \leq v_n \leq \frac{n}{4\pi} H(\delta) [\int D_2(t) d\Phi(t)]^2$$

where

$$H(\delta) = \sum_1^p \exp(-n\delta_i^2/2) + \left(\sum_1^p \exp(-n\delta_i^2/4) \right)^2.$$

The main point is that under certain circumstances we may approximate v_n by $n H(\delta)/4\pi$, which is small for large n . This holds when the dispersion of the δ -values is not too large, so that $\delta_1 < 2\delta^*$. In this case both integrals above are close to 1 as $n \rightarrow \infty$.

Finally, we note that $H(\delta)$ is Schur-convex in $(\delta_1^2, \dots, \delta_p^2)$, so that further bounds on H may readily be obtained. In particular, it can be shown via such bounds that v_n decreases exponentially in n , so that we can expect reasonably good approximations for moderate values of n .

Table 1. Numerical Values of $\int_{-\infty}^{\infty} \prod_{i=1}^p \psi(y+id) d\psi(y)$
for $d = 0.00(0.02)0.20(0.10)4.00$, $p = 2(1)10, 12, 15, 20$

<u>d</u>	<u>P</u>	2	3	4	5	6	7	8	9	10	12	15	20
0.00	.3333	.2500	.2000	.1667	.1429	.1250	.1111	.1000	.0909	.0769	.0625	.0476	
0.02	.3418	.2604	.2118	.1795	.1567	.1396	.1264	.1159	.1074	.0945	.0814	.0683	
0.04	.3504	.2709	.2238	.1928	.1710	.1549	.1425	.1329	.1251	.1135	.1023	.0920	
0.06	.3589	.2815	.2360	.2063	.1857	.1707	.1593	.1506	.1437	.1337	.1246	.1172	
0.08	.3676	.2922	.2484	.2202	.2008	.1870	.1767	.1689	.1630	.1547	.1477	.1429	
0.10	.3762	.3030	.2610	.2343	.2163	.2036	.1945	.1877	.1827	.1761	.1710	.1681	
0.12	.3849	.3139	.2737	.2486	.2319	.2205	.2125	.2068	.2027	.1975	.1940	.1924	
0.14	.3936	.3249	.2865	.2630	.2478	.2376	.2307	.2259	.2226	.2188	.2165	.2157	
0.16	.4024	.3360	.2994	.2775	.2637	.2547	.2489	.2450	.2424	.2396	.2382	.2378	
0.18	.4111	.3470	.3124	.2921	.2797	.2719	.2669	.2638	.2619	.2600	.2591	.2589	
0.20	.4199	.3581	.3254	.3067	.2956	.2889	.2848	.2824	.2810	.2797	.2792	.2791	
0.30	.4636	.4136	.3901	.3786	.3730	.3704	.3692	.3687	.3684	.3683	.3683	.3683	
0.40	.5067	.4679	.4524	.4462	.4439	.4432	.4429	.4428	.4428	.4428	.4428	.4428	
0.50	.5487	.5199	.5104	.5075	.5068	.5066	.5066	.5065	.5065	.5065	.5065	.5065	
0.60	.5892	.5686	.5633	.5621	.5619	.5619	.5619	.5619	.5619	.5619	.5619	.5619	
0.70	.6301	.6164	.6138	.6134	.6134	.6134	.6134	.6134	.6134	.6134	.6134	.6134	
0.80	.6638	.6546	.6533	.6532	.6532	.6532	.6532	.6532	.6532	.6532	.6532	.6532	
0.90	.6976	.6918	.6912	.6912	.6912	.6912	.6912	.6912	.6912	.6912	.6912	.6912	
1.00	.7288	.7253	.7251	.7251	.7251	.7251	.7251	.7251	.7251	.7251	.7251	.7251	
1.10	.7574	.7574	.7573	.7573	.7573	.7573	.7573	.7573	.7573	.7573	.7573	.7573	
1.20	.7836	.7825	.7825	.7825	.7825	.7825	.7825	.7825	.7825	.7825	.7825	.7825	
1.30	.8074	.8068	.8068	.8068	.8068	.8068	.8068	.8068	.8068	.8068	.8068	.8068	
1.40	.8290	.8287	.8287	.8287	.8287	.8287	.8287	.8287	.8287	.8287	.8287	.8287	
1.50	.8485	.8483	.8483	.8483	.8483	.8483	.8483	.8483	.8483	.8483	.8483	.8483	
1.60	.8661	.8660	.8660	.8660	.8660	.8660	.8660	.8660	.8660	.8660	.8660	.8660	

References

- [1] Bechhofer, R.E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Statist., 25, 16-39.
- [2] Gupta, S.S. (1963). Probability integrals of multivariate normal and multivariate t. Ann. Math. Statist., 34, 792-828.
- [3] Hardy, G.H., Littlewood, J.E. and Polya, G. (1934, 1952). Inequalities. 2nd ed. Cambridge University Press.
- [4] Kimball, A.W. (1951). On dependent tests of significance in the analysis of variance. Ann. Math. Statist., 22, 600-602.
- [5] Marshall, A.W. and Olkin, I. (1974). Majorization in multivariate distributions. Ann. Statist., 2, 1189-1200.
- [6] Milton, R.C. (1970). Rank order Probability: Two sample shift alternatives. John Wiley and Sons, New York.
- [7] Milton, R.C. (1972). Computer evaluation of one multivariate normal integral. Technometrics, 14, 881-889.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE			READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER <i>TR-174</i>	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) <i>ESTIMATING THE TRUE PROBABILITY OF CORRECT SELECTION FOR LOCATION AND SCALE PARAMETER FAMILIES</i>			5. TYPE OF REPORT & PERIOD COVERED <i>Technical Report</i>
6. AUTHOR(s) <i>I. Olkin, M. Sobel Y.L. Tong</i>			7. PERFORMING ORG. REPORT NUMBER <i>N00014-75-C-0561 NSF-MPS-75-0745A</i>
8. PERFORMING ORGANIZATION NAME AND ADDRESS Dept. of Operations and Dept. of Statistics Stanford University Stanford, California 94305			9. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS <i>NR-042-022</i>
10. CONTROLLING OFFICE NAME AND ADDRESS Statistics & Probability Program Code 436 Office of Naval Research Arlington, Virginia 22217			11. REPORT DATE <i>14/28 June 1976</i>
12. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) <i>(12) 21 P.</i>			13. NUMBER OF PAGES <i>17</i>
14. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED.			15. SECURITY CLASS. (of this report) <i>UNCLASSIFIED</i>
16. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)			17a. DECLASSIFICATION/DOWNGRADING SCHEDULE
18. SUPPLEMENTARY NOTES This report was also supported by National Science Foundation Grant MPS 7509450-000			
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) MULTIPLE COMPARISONS RANKING & SELECTION MAJORIZATION PROBABILITY OF CORRECT SELECTION			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The usual formulation of the problem of selecting the best of k populations has been studied from the point of view of designing an experiment that will guarantee a certain probability of making a correct selection (PCS). However, in retrospective studies the sample has already been taken and we wish to estimate the probability of making a correct selection using this data. The present study deals with obtaining a class of upper and lower bounds for the true PCS and its estimator using the methods of majorization. We also study the large sample behavior of the estimator of the PCS, which turns out to be asymptotically normal.			

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 68 IS OBSOLETE
S/N 0102-014-6601UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

402 766 mt

